DOCUMENT RESUME

ED 190 489 SP 016 201

AUTHOR Fox, Robert A.
TITLE How to Design, Construct and Evaluate a Multiple-Choice Examination in Health Education--A Practical Approach.
PUB DATE [79]
NOTE 11p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education: *Health Education: *Multiple Choice Tests: *Test Construction: Testing Problems: *Test Items: Test Validity

ABSTRACT

A well developed multiple choice test is a reliable instrument for grading students and evaluating teacher presentation. There are three steps in the development of a valid multiple choice examination: 1) design or "blueprinting," 2) item construction, and 3) item analysis and evaluation. "Blueprinting" is the identification of the types of knowledge to be measured and the relative importance of each. The actual writing of the questions, or item construction, follows. Item evaluation examines each question to test its validity and clarity. Explanations and examples for each step are included in this guide. (CJ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED190489

HOW TO DESIGN, CONSTRUCT AND EVALUATE A MULTIPLE-CHOICE EXAMINATION IN HEALTH EDUCATION - A PRACTICAL APPROACH

Dr. Robert A. Fox

Assistant Professor

Department of Health and Physical Education

John Jay College of Criminal Justice

The City University of New York

444 West 56th Street

New York, N.Y. 10019

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Robert A. Fox

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SP 016 201

HOW TO DESIGN, CONSTRUCT AND EVALUATE A MULTIPLE-CHOICE EXAMINATION IN HEALTH EDUCATION - A PRACTICAL APPROACH

Virtually every teacher at every level is required to measure the achievement of his or her students. This is often difficult because the skills involved in designing, constructing and evaluating testing instruments are seldom included in the baccalaureate and masters level teacher education curricula. It is not surprising that teachers often lack confidence in their testing procedures and that students feel anxious before taking teacher-made tests. Student uncertainty exists because too often, teachers unskilled in test preparation test trivial pieces of information or concepts of particular interest to themselves rather than a representative sample of the essential material. This situation is unfortunate because the basic skills necessary for reliable and valid achievement testing are easily acquired.

The purpose of this article is to present a practical approach to the acquisition of the skills necessary to overcome these common pitfalls. The focus will be on the multiple-choice achievement examination, the main purpose of which is to provide a reliable instrument for grading students. At the same time, the instructor will be able to ascertain some measure of how effectively the material has been presented.

There are three steps in the development of a reliable and valid multiple choice examination:

1. Design or "Blueprinting"
2. Item Construction
3. Item Analysis and Evaluation

The Design

The first step in designing a multiple-choice examination is to "blueprint" the test. This is done by analyzing what is to be measured, and in most cases this corresponds to the major objectives of the unit. For example, a typical nutrition unit might "blueprint" as follows:

Topic	Comprehension	Application	Total
The Nutrients	10%	20%	30%
Digestion	10	10	20
Metabolism	10	10	20
Diet and Weight Control	10	20	30
	40%	60%	100%

The objectives in this unit have been divided into four major topic areas: the nutrients, digestion, metabolism and diet and weight control. A percentage or weight has been assigned to each topic area corresponding to its importance and the emphasis it received in class. In this case, the nutrients received 30%, digestion 20%, metabolism 20% and diet and weight control 30%. The weight or percentage indicates the total number of questions on the test for each topic area. The topic areas are further delineated into "process" areas. These "process" areas change with the subject matter and level of the class and relate to the categories of the cognitive domain: knowledge, comprehension, application, analysis, synthesis and evaluation. In the example cited, the "process" areas are comprehension and application.

In the sample "blueprint" 30% of the items will test the nutrients; 10% comprehension and 20% application. On a 50 item test, five questions (10%) would involve the comprehension of

material involving nutrients and 10 items (20%) would correspond to the application of information on nutrients. The fifty item examination is created by proceeding through each topic area in the same fashion.

It is important to construct items that measure the understanding of key concepts. It is equally important that the test items collectively represent an instrument that will discriminate between students who have learned the material and those who have not. A "hard" or "easy" test is meaningless. Only a test that successfully teases apart those who have learned from those who have not is reliable and valid.

Item Construction

The next step is the actual writing of the questions or item construction. First of all, the directions to the test must be explicit and must include how the test is to be graded. For example: "Choose the best answer to each question. Only one answer is correct for each item. Answer all questions because your score is based on the number of correct answers."

In this example the topic area "the nutrients" will encompass 30% of the test. Within that topic area application of the concepts regarding "the nutrients" will comprise 20% of the total test items. If one of the unit objectives was "the student will be able to identify foods in terms of their nutritive values," an application items might be:

Meat and fish are good sources of

a. carbohydrate
b. fat
c. fiber
d. protein

It takes time to design items that reliably correspond to the objectives, the "blueprint" and the level of the students. However, with practice, the construction of good items becomes easier and less time-consuming. Also, the teacher can store all questions in a test bank for future use.

There are two parts to the multiple choice item: the stem, which introduces the question, and the alternatives or options. The incorrect options are called distractors.

The stem should include only as much information as is necessary to answer the item. The stems should be unambiguous and the reading level as simple as possible. Words with precise meaning should be employed and imprecise words such as fair and pretty good should be avoided. Try to avoid negative statements. However, if you anticipate the possibility of confusion underline a word or phrase for clarity. For example: Each of the following foods contains protein <u>except</u>

<u>a</u>. cabbage
b. fish
c. meat
d. poultry

Do not allow one item to give a clue to the answer of another. Items should be mutually exclusive. Include all pertinent information in the stem and keep the options as short as possible. An example of an item in which too much information is contained in the options is:

Meat and fish

a. are good sources of carbohydrate
b. are good sources of fat
c. are good sources of fiber
<u>d</u>. are good sources of protein

It is better to ask: Meat and fish are good sources of

a. carbohydrates c. fiber

b. fat d. protein

The distractors or incorrect options must be grammatically consistent with the stem. For example: Meat and fish are good sources of a

a. carbohydrates c. fiber

b. fats d. protein

Note that options a, b and c are in the plural leaving d as the only option grammatically consistent with the stem.

Distractors should consistently offer either three or four reasonable alternatives of about the same length. One answer should be clearly best or correct. All of the above, none of the above, a and b and so forth are confusing and should be avoided. Distractors should not include specific determiners like never, all, always and every because they are usually wrong.

Use a table of random numbers to randomize the order of correct answers so there is no pattern. It is not a good practice to have equal numbers of a, b, c and d as correct answers.

The skill of writing multiple choice test items comes with practice. When an item seems to be particularly difficult to word it sometimes helps to reverse the information from the stem to the alternatives or vice versa. For example, instead of asking: "Which of the following food groups is a good source of protein?"

a. breads and cereal c. fruits and vegetables

b. dairy d. meat and fish

the question might read: "Meat and fish are good sources of

a. carbohydrate c. fiber

b. fat d. protein

In deciding on specific material to be used in item construction, the teacher should attempt to test understanding of important concepts rather than recall or recognition of such things as names, dates and other trivia. A common weakness of teacher-made tests is the focus on rote memory of classroom or textbook material. It is good technique to use novel situations to test the application of important concepts. For example:

Rote memory: Which of the following foods contains the most calories per gram?

a. fat
b. protein
c. carbohydrate
d. sugar

Novel situation: Peter wishes to lose weight. Which of the following would result in the greatest reduction of calories if it were eliminated from his daily diet?

a. 1 tablespoon of butter
b. 1 tablespoon of granulated sugar
c. 1 slice of white, enriched bread
d. 1 boiled egg

Do not use verbatim statements from textbooks. Trick or "catch" questions are not good because an achievement test should measure knowledge, not I.Q. or "test-savvy."

Most exams should not be for time unless they are power tests. A good method for determining sufficient time is for the teacher to take the exam and multiply that time by four.

Item Analysis and Evaluation

There are several systems for evaluating multiple-choice exams; the one which follows is effective yet simple.

Set up a chart as shown below for each test item. If 100 students took the test used the top 25 and lowest 25 scores. The figures indicate the number of students in each percentile who chose a, b, c or d as an answer. For item #1, c was the correct answer.

		a	b	c	d
Item #1	Upper 25%	1	2	20	2
	Lowest 25%	7	6	5	7

One good measure of reliability is the ability of an item to effectively discriminate between better and poorer students. Item #1 is a reliable item because 20 of the better students got it correct while only five of the poorer students did. This indicates that the item did a good job of discriminating between the better and poorer students. The greater the difference between these two groups the more reliable the item. Also note that the incorrect alternatives or distractors a, b and d received about equal attention as answers. This indicates good distractors. Had no one or few students selected a particular distractor it might be an indication to replace that distractor with a more attractive one if the item is to be reused.

In this next example, b is the correct answer.

		a	B	c	d
Item #2	Upper 25%	0	10	15	0
	Lowest 25%	6	10	2	7

Item #2 is a poor discriminator and, therefore, an unreliable item because it did not discriminate between the better and poorer students; an equal number from each group (10) selected b as the correct answer. In addition, 15 of the better students opted for distractor c. This is a sure sign that something is wrong. Why

were so many better students attracted to this distractor? Also note that only two of the poorer students selected c.

In summary, examine each item to see that it discriminates between better and poorer students and provides distractors that are reasonably attractive especially to poorer students. Sometimes an item can be improved by changing a distractor because it is attracting too many of the better students or not enough of the poorer ones.

The overall difficulty of an item is the percentage of total students who got the item correct. Sometimes an item of low difficulty or an easy question is used to ensure that a concept was learned. More difficult items tend to be more reliable in that they discriminate better between more and less knowledgeable students. A teacher can average items from past exams in order to obtain the desired level of difficulty. Many measurement experts recommend a 70% difficulty level.

It is good educational procedure to grade and return a test to the class the first meeting after it is taken. This gives the student an opportunity to discuss the exam while the experience is fresh in his or her mind. Each question should be reviewed with the correct answer discussed. Students should be encouraged to express any confusion they had with a question. If several students express confusion with an item it is a good idea to eliminate or revise that question.

BIBLIOGRAPHY

Bloom, B. Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain. New York: David McKay Company, Inc., 1956.

Mehrens, W. and Lehman I. Measurement and Evaluation in Education and Psychology (2 ed.). New York: Holt, Rinehart and Winston, Inc., 1978.

Noll, V., Scannell, D. and Craig R. Introduction to Educational Measurement (4 ed.). Boston: Houghton Mifflin Company, 1979.

Stanley, J. and Hopkins, K. Educational and Psychological Measurement and Evaluation (5 ed.). New Jersey: Prentice-Hall, Inc., 1972.

Thorndike, R. and Hagen E. Measurement and Evaluation in Psychology and Education (4 ed.). New York: John Wiley and Sons, Inc., 1979.